

CLAIMS

- 1 1. A method for determining a number of future content
2 requests that will arrive at an information delivery system for a pre-
3 determined future period of time, comprising:
4 creating a plurality of models to predict a number of future content
5 requests;
6 determine for each a model its respective prediction for the pre-
7 determined future period of time;
8 selecting a model from the plurality of models which has a least
9 error associated with its prediction to create a best model predictive
10 assessment of the next interval's number of content requests;
11 adding the number of current content requests with the predicted
12 future content requests to create an aggregate total number of content
13 requests; and
14 sending the aggregated total number of content requests to a capacity
15 function.
- 1 2. The method of claim 1, wherein the least error is
2 (a measured number of content requests- a predicted number of content
3 requests)².
- 1 3. The method of claim 1, wherein the least error is a method to
2 determine accuracy of a model that predicts the number of content requests.
- 1 4. The method of claim 1, wherein the least error is determined
2 by observing of the number of content requests during a selected time
3 period and then comparing the number of content requests observed with a
4 predicted number of content requests.

1 5. The method of claim 1, wherein the least error is determined
2 at an instant period of time.

1 6. The method of claim 1, wherein the least error is determined
2 over a period of time.

1 7. The method of claim 1, wherein the least error changes with
2 modifications of a user's quality of service objectives.

1 8. The method of claim 1, wherein the least error changes with
2 modifications to the information system.

1 9. The method of claim 1, wherein selecting the model includes
2 construction of a probability distribution over a set of predictive models.

1 10. The method of claim 9, wherein construction of the
2 probability distribution determines the accuracy of the plurality of models
3 and a stochastic selection of the plurality of models according to the
4 probability distribution.

1 11. A method for determining a number of future content
2 requests that will arrive at an information delivery system for a pre-
3 determined future period of time, comprising:
4 receiving a user's quality of service objectives at the information
5 system;
6 creating a plurality of models to predict a number of future content
7 requests;
8 determine for each a model its respective prediction for the pre-
9 determined future period of time;

10 selecting a model from the plurality of models which has a least
11 error associated with its prediction to create a best model predictive
12 assessment of the next interval's number of content requests;
13 adding the number of current content requests with the predicted
14 future content requests to create an aggregate total number of content
15 requests;
16 sending the aggregated total number of content requests to a capacity
17 function;
18 determining if a content request is for an existing session or a new
19 session; and
20 sending the content request to a dispatch control function at the
21 information system when the content request is for an existing session.

1 12. The method of claim 11, wherein the user's quality of service
2 objectives include speed of content delivery for a specified time.

1 13. The method of claim 11, wherein the user's quality of service
2 objectives include consistency of speed of content delivery.

1 14. The method of claim 11, wherein the user's quality of service
2 objectives include a function of number of concurrent users.

1 15. The method of claim 11, wherein the user's quality of service
2 objectives include system response time.

1 16. The method of claim 11, wherein the user's quality of service
2 objectives include system response time consistency.